LEVEL III

RESEARCH ON THE TECHNOLOGY OF INFERENCE AND DECISION

SOCIAL SCIENCE RESEARCH INSTITUTE
UNIVERSITY OF SOUTHERN CALIFORNIA

A048567

Ward Edwards
Richard John
William Stillwell

ADVANCED ARPA DECISION TECHNOLOGY PROGRAM

CYBERNETICS TECHNOLOGY OFFICE
DEFENSE ADVANCED RESEARCH PROJECTS AGENCY
Office of Naval Research • Engineering Psychology Programs

The objective of the Advanced Decision Technology Program is to develop and transfer to users in the Department of Defense advanced management technologies for decision making. These technologies are based upon research in the areas of decision analysis, the behavioral sciences and interactive computer graphics. The program is sponsored by the Cybernetics Technology Office of the Defense Advanced Research Projects Agency and technical progress is monitored by the Office of Naval Research — Engineering Psychology Programs. Participants in the program are:

Decisions and Designs, Incorporated
Harvard University
Perceptronics, Incorporated
Stanford Research Institute
Stanford University
The University of Southern California

Inquiries and comments with regard to this report should be addressed to:

Dr. Martin A. Tolcott
Director, Engineering Psychology Programs
Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217
or
LT COL Roy M. Gulick, USMC
Cybernetics Technology Office
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, Virginia 22209

11P

TECHNICAL REPORT SSRI 77-6

6

RESEARCH ON THE TECHNOLOGY OF INFERENCE AND DECISION.

by

10
Ward Edwards, Richard John [redacted] William Stillwell

Sponsored by

Defense Advanced Research Projects Agency
ARPA Order ~~No.~~ 3052
Under Subcontract from
Decisions and Designs, Incorporated

15 N00014-76-C-0074

11 November 1977

9 Final rept.
1 Oct 76-30 Sep 77

12 39p.

SOCIAL SCIENCE RESEARCH INSTITUTE
University of Southern California
Los Angeles, California 90007

Contents

Acknowledgments

This research was supported by the Advanced Research Projects Agency of the Department of Defense and was Monitored by the Office of Naval Research under Contract N00014-76-C-0074 under subcontract from Decisions and Designs, Inc.

The authors would like to thank David Seaver, Lee Eils, Tsuneko Fujii, William Gabrielli, Peter Gardiner, Patrick Leung, and J. Robert Newman, who made many valuable contributions to the research summarized in this report.

Disclaimer

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily respresenting the official policies, either expressed or implied, of the Advanced Research Projects Agency or of the United States Government.

Introduction

This Final Report summarizes the work by the Social Science Research Institute, University of Southern California on subcontract P.O. 76-030-0715 from Decisions and Designs, Inc., prime contract N00014-76-C-0074 from the Advanced Research Projects Agency, monitored by the Engineering Psychology Programs, Office of Naval Research. The research conducted during this contract period from October 1, 1976 to September 30, 1977 under the direction of Professor Ward Edwards, the Principal Investigator, was part of an ongoing program of Research on the Technology of Inference and Decision. Edwards (1973, 1975) and Edwards and Seaver (1976) summarized previous research.

The proposal leading to this subcontract called for research on five specific topics: measurement and validation of multiattribute utilities, sensitivity to approximation of multiattribute models and assessment procedures, group processes for probability assessment, assessing very small probabilities, and biases in subjective probability distributions on non-percentage variables. Our research on these and other topics is reported in ten technical reports which have been produced or are now being prepared. Summaries of these technical reports appear at the end of this report.

The purpose of this report is to explain how this research integrates into an overall program of research on decision technology. Thus, we do not report in detail findings that are set forth in the self-contained technical reports. Only major findings are reviewed along with ongoing research and future research possibilities suggested by our current work.

II. A Technical Overview

Research at SSRI has sought to determine the strengths which decision makers bring to the decision situation as well as to study decision aids and techniques that improve reliability and validity of judgments. Both theoretical and practical topics have been examined; the line between today's theoretical research and tomorrow's tool or technique having been deliberately blurred. In the practical vein, we have studied simplification techniques for construction of utility models, aids to the decision maker in the assessment of small probabilities, effects of various response modes for the elicitation of probabilities, and comparision of group behavioral techniques versus mathematical aggregation models for group probability assessment. More theoretical work has sought to determine the individual's ability to deduce distributions underlying the generation of stimuli. Recently we have become interested in both qualitive and quantitative aspects of expertise as it affects the judgment of uncertainty.

II A. Elicitation and Quantification of Uncertainty

II A.1 Group assessment of uncertainty: Human interaction versus mathematical models.

Often a decision maker is not a single individual but rather several, each of whom should be able to influence the final decision. In decision analysis a single judgment of uncertainty as well as a single judgment of value or utility is necessary as input to each branch of the decision making structure. This apparent incompatibility has led to much research into techniques whose aim is to derive from the group a single value for each measure of uncertainty. Such research has explored two major strategies, mathematical techniques for the aggregation of individual judgments into a single group estimate, and behavioral techniques which seek group consensus.

Either approach has both mathematical and social psychological difficulties. Dalkey (1972) has shown that no formal rule for the aggregation of individual probabilities can satisfy a set of reasonable conditions (such as non-dominance by a single group member). A comparable proof exists for utility judgments (Arrow, 1951). Behavioral techniques likewise have limitations. Individual group members may concern themselves more with reaching consensus than with the quality of the agreed-on judgment. Various factors such as individual dominance through personality characteristics or rank within the organization may influence judgments despite their irrelevance to the task.

In an effort to compare various behavioral and mathematical techniques of group probability assessment, Seaver (1977, in press) experimentally compared two aggregation rules, weighted arithmetic means and weighted geometric means, and three weighting procedures, equal weights, weights based on self-rating and DeGroot weights (DeGroot, 1974). Five behavioral interaction techniques were compared, the Delphi method (Dalkey and Helmer, 1963) the Nominal Group Technique, developed by Delbecq and Van de Ven (1971), a modified nominal group technique in which group members state their estimates and reasons with no discussion, a consensus technique in which groups were to arrive at consensus in any way they wished, and a no interaction or control group in which group members made estimates with no knowledge of other group members' estimates.

The quadratic scoring rule was used as the criterion for measuring the quality of group assessments. The well-known insensitivity of that rule may account for the lack of significant differences among behavioral techniques. In general, interaction among group members reduced differences, reduced the calibration of the judgments, and increased the extremeness of judgments. Therefore, deciding whether or not to use group interaction techniques involves a tradeoff between calibration and extremeness of responses. Although no significant differences were found, slight differences as well as the results of other studies point to slight superiority of the nominal group technique to other group interaction methods.

The data show that little if anything is lost by using mathematical techniques to aggregate individual judgments rather than behavioral interaction. Considering the practical disadvantages of face-to-face meetings of groups, no point exists in bothering with the sometimes lengthy procedures of behavioral interaction. While results of this experiment dealt totally with point estimates, further studies will attempt to elicit continuous distributions. Results of these studies will become available over the next several months.

II A.2. Response scale effects on likelihood ratio judgment.

Several studies, (Goodman, 1973; Phillips and Edwards, 1966) have shown consistent effects for different methods of elicitation of subjective judgment. Stillwell, Seaver and Edwards (1977) defined the effects due to the scale on which the subjects responded but at the time it was felt that results could at least partly be due to the extreme diagnosticity of the data to which the subjects responded. A d' of 3.0 was used to generate data and the range of veridical likelihood ratios to which the subjects responded was such that likelihood ratios as high as 12,000:1 were encountered. Although significantly better (closer to veridical) responses were found when the subjects were responding to logarithmically spaced scales, it was felt that a more moderate d' and range of likelihood ratios might reduce the magnitude of differences or change the relationship altogether. A second experiment was therefore undertaken in which d' and the range of true likelihood ratios were varied.

The results resembled those of the first experiment. Logarithmically spaced scales were superior to linearly spaced scales. The range of true likelihood ratios, was, however, shown to have a strong and significant effect on performance. Subjects were much better able to approximate veridical judgments when less extreme true likelihood ratios were chosen. There was also a significant interaction between endpoint and spacing (logarithmic versus linear) accounting for a relatively large proportion of the variance.

Scale endpoints were shown to influence judgments consistently. Either of two factors may be contributing to this finding. It is possible that the upper endpoint offers an upper bound to responses thereby limiting the range of values expressed. A second possibility is that the endpoints controlled subjects' judgments about the range in which they could expect the true value to fall. More extreme endpoints may thus produce more extreme responses. For a more detailed description of the results of this study, see summary No. 2.

II A.3. Averaging as a means of probabilistic inference.

Edwards and Seaver (1976) discussed an experiment by Eils, Seaver and Edwards (1977) in which averaged log likelihood ratios were elicited from subjects and used as inputs to a probabilistic information processing (PIP) system. That is, such log likelihood ratios judgmentally averaged over all data and then processed by means of Bayes' theorem produced more extreme final odds than posterior odds estimated directly. In light of the general finding of conservatism in probability revision tasks, this would suggest that PIP outputs are more likely to reflect subjective certainty than are posterior odds judgments. Experiment I also showed that persons using the averaged log likelihood ratio judgements were more orderly in these judgments as evidenced by higher correlation between true final odds and final odds calculated via Bayes' theorem.

A second experiment was undertaken in order to determine whether the judgmentally averaged log likelihood ratio technique contributed significant improvement over the likelihood ratio judgment originally proposed for the PIP system developed by Edwards et al (1968). Also a variable in Experiment II was the diagnosticity of the data used to elicit subjects responses. It was found that data diagnosticity affected quality of response for both response modes. Estimates became more veridical as the data became more diagnostic. The primary finding of the study was that quality of estimates did not differ significantly in either veridicality or orderliness between likelihood ratio estimates as originally proposed for the PIP technique and for the averaged log likelihood estimates. Both methods were found to produce better estimates than cumulative certainty judgment, as is usual in such comparisons.

The reason for considering an alternative to likelihood judments is that a problem may arise in applying PIP systems in real world contexts. The people assessing the likelihood ratios will typically have access to feedback about the posterior odds that are calculated from their likelihood ratios. Goodman (1973), in a reanalysis of data from five studies exploring methods of eliciting judgments about uncertain events, concludes that feedback about the implications of judgments makes them less extreme and is probably the most powerful variable controlling the extremeness of the judgments. Thus, even a PIP system may be susceptible to conservatism in real world applications. This problem seems less likely to characterize judgments of average certainty due to the very nature of the elicited judgments. Should further research confirm feedback produced conservatism in PIP systems, average certainty judgments may prove to be a useful alternative to PIP.

These findings also compel a rethinking of the misaggregation explanation of conservatism in probability revision. Mean log likelihood ratio is a judgmentally aggregated response--but it is not conservative (nor yet radical). Apparently, aggregation that has the character of a sum or product (i.e. the target number is outside the range of input quantitities) is conservative. Aggregation that has the character of an average (the target number lies near the middle of the range of input quantities) is unbiased.

II. A.4 The assessment of small probabilities.

In the study of small probability, high (either positive or negative) expected value decision making situations, decision analysis can make significant contributions. Nuclear engineering has brought this situation to public attention as system failures occur with probabilities typically smaller than 10^{-6} but with values which may exceed 50,000 lives lost. But identical kinds of problems arise frequently in military and political contexts. An obvious example is whether or not a particular limited-war strategy may lead to a widening of the war. Both experimental and applied work have shown, however, that problems arise in the subjective assessment of the likelihood of highly unlikely events.

Unpublished work by Slovic, Lichtenstein, Fischoff, Coombs, and Layman suggests a remedy for the small probability assessment problems.

Instead of direct assessment of the probability of interest, Slovic, et al. asked subjects to judge which of two events was the most likely. They found, with a few notable exceptions, that over eighty percent of subjects could correctly judge the larger of the probabilities of a pair of events when the ratio of the probabilities was greater than 2:1.

These findings suggest that either a series of comparisions of event pairs or a simultaneous comparison of the event with unknown probability with a list of events with known probability may result in significant improvement in probabilistic judgments. Several studies were undertaken to evaluate the potential of this approach. In the first set of these experiments subjects were asked to place the event of interest into a list of events at a point appropriated to its relative likelihood of occurrence. Incentive was given to subjects in the form of $3.00 for each response placed in the correct space among thirty spaces between events. Results of this experiment show that subjects are not sufficiently able to perform the task to warrant the techniques used as an elicitation tool for probabilities. The mean correlation over 120 subjects between response probability (using the midpoint of the response space) and the true probability (using the midpoint of the space in which the event should be placed) was .131.

Because the above results might be due to the cognitive difficulty of simultaneously comparing an event with thirty-one other events, a second experiment asked subjects to compare the event of interest with either one or two other events. A branching structure was used whereby the subject, after making a series of judgments, would arrive at an estimate of the probability of the response event. This probability was evaluated in the same manner as in the previous experiment using the midpoint of the space arrived at by the branching process. Results of this experiment are comparable with those of the first experiment. Correlations are, on the average, slightly positive but for no subject were they significant or large enough to justify the technique.

The findings of these two experiments raise a significant question. How does one explain the apparent divergence of these results from those of Slovic et al? The nature of the events used in the current series of experiments differed from those used by Slovic for three-fourths of the subjects; but analysis of the one-fourth using occupation events,

selected from those used by the Slovic group, shows the same inability to make accurate judgments. Closer examination of the task explains the discrepancy. While subjects in the Slovic et al. experiment produced relatively good directional judgments when the odds ratio between events was greater than 2:1, event pairs with odds ratios less than 2:1 produced very poor judgments. Frequently, subjects systematically choose the less probable event as more likely. In the current experiments, subjects were required to make successively more sensitive judgments so that even if their initial pairwise judgments were correct, later judgments of the branching or list placement tasks involved choices too sensitive for their abilities.

We have two ideas about the usefulness of techniques developed out of the work of Slovic et al. First, for the skill of the probabilistic judge to be effective in realizing improved estimates it seems likely that aggregation over individuals must occur in some form. Slovic found that directional choices in the pairwise task were likely to be correct for a large proportion of subjects, but not all persons were correct and most odds ratio judgments were too conservative. A sort of majority rule principle in paired comparisions of probabilities may well yield improved probabilistic judgments.

A second issue concerns the nature of expertise. A serious question for the use of decision analysis is and has been: Can experts make the needed probabilistic judgments? Experimentation on probabilistic bias suggests that these judgments are of decidedly low quality although there is evidence that experts perform these tasks somewhat better than college sophomores, such as we have been using. A pilot study has been performed which points to some interesting possibilities in this area.

We defined expertise as experience or familiarity with the subject matter whose relative likelihood was to be judged. In the case of the initial experiment this was baseball statistics for the Los Angeles Dodgers players. Subjects were asked ten questions of the type, "Is it more likely that a randomly selected Dodger player (not a pitcher) has 25 or more home runs this season or is it more likely that a randomly selected Dodger player (not a pitcher) has 88 or more bases on balls this season?" At the same time subjects were asked ten questions about

the relative likelihood that a person selected at random would turn out to be, for example, a lawyer or secretary. Those questions about employment were taken directly from the Slovic et al. study so that a comparison between subjects could be made.

The pattern of responses for the Slovic et al. questions in this study was roughly the same as it was in the original study. Percentages of correct directional responses were slightly higher for six questions and lower for the other four than for Slovic et al. Our subjects were therefore comparable to those in the original study.

As a measure of expertise subjects answered a series of questions about themselves concerning the number of games they had attended, how many times they read the box scores, etc., as well as the number of rostered Dodger players they could name. They also rated themselves as Dodger fans on a 7-point scale. Regression analysis was then done with these measures of baseball expertise as predictors of quality of performance on both the Dodger questions and the Slovic et al. questions.

The multiple regression results are very similar for Dodger and Slovic et al. questions. The multiple regression coefficients were .55 and .48 respectively. That is, Dodger fans do better on questions about employment, as well as about the Dodgers, than non-Dodger fans. The similarity between these two coefficients suggests the possibility that a common factor underlies the ability to answer both types of questions. Perhaps expertise in a given subject area is not the important factor in performance of a probabilistic task. Maybe the ability to deal with probabilistic thought, and thereby put whatever substantive knowledge is available to use, is what produces good probabilistic assessments.

This though is not original with us. Winkler (1967) discusses two forms of expertise, one in which substantive information is brought to the task and a second in which the subject understands the probabilistic task and the concept of uncertainty. Further study of the regression analysis results shows similar patterns for the beta weights on Dodger and Slovic et al. questions. Therefore, the pattern of information used in the two different types of questions is remarkably similar. This finding is congruent with the two forms of expertise hypothesis.

The results of this line of experimentation have raised many more questions than they have answered. Study into the nature of expertise will obviously be a fruitful line of endeavor and should precede further work on marker event techniques. It could well be that training of experts in probabilistic thinking would lead to significant improvement in the quantification of uncertainty.

II A.5. Estimating subjective probability distributions.

Probably the most cited upublished work in decision theory, that of Alpert and Raiffa (1969), found that different methods for assessing proability distributions resulted in different levels of bias in responses, in their case "too tight distributions". Seaver, von Winterfeldt and Edwards (1975) went on to show that the amount of bias as measured by "surprises", a true value falling outside a specific central interval of the assessed distribution, was affected in systematic ways by the method used to assess the probability distribution. These results are consistent with many others in decision theoretic research in which assessment techniques have contributed to the quality of probabilistic judgment.

One problem with studies of probability assessment is that the available dependent measures are not completely satisfying ones. Proper scoring rules do not provide a sensitive measure of how closely an elicited probability distribution reflects the assessor's beliefs about the chances of occurrence of the events over which the distribution is assessed. Typical measures of calibration involve the proportion of events which, historically and across distributions, lie on fixed intervales of the assessed distribution. The proportion of events obtained that were assessed as lying within the interquartile interval (.25 p .75) is an example. Such measures require assessments

over huge numbers of distributions to be reliable. Furthermore, many interesting questions regarding the usefulness of elicitation techniques simply cannot be answered without a measure which takes into account a larger number of the important features which distinguish one probability distribution function from another.

John & Edwards (1977) investigated the possibility of presenting Ss with a sample distribution of a random variable and eliciting the population distribution (density) from which the sample was presumably drawn. Stimuli were pickup sticks (length = 6.5 inches), painted blue and yellow. The length of yellow on each stick constituted the random variable. Subjects were shown three sample distributions (uniform, modal, and bimodal) of twenty-six sticks each.

Each subject used one of three probability elicitation procedures to convey his (her) knowledge of the population distribution from which the sample was presumably drawn. In the fractile procedure, Ss were asked "to give a length of yellow such that a stick chosen randomly from the population just sampled will have a length of yellow less than or equal to the length you give with probability = (.99,.75,.50,.25,.01)". In the probability procedure, Ss were asked "to judge what the probability is that a stick chosen randomly from the population just sampled has a length of yellow less than or equal to (.65",1.95", 3.25", 4.55", 5.85")". A third procedure (graph) required Ss to draw a curve, of which "the height at each point represents the relative probability that a stick drawn at random from the population will have that length of yellow". The fractile and probability methods were used by Seaver, von Winterfeldt, and Edwards (1975) and essentially involve obtaining estimates of five points (ordered pairs) along the cumulative distribution. The graph technique obtains a sketch of the S's density function.

For each of the curves produced using the Graph technique, fourteen points evenly spaced along the curve were used as input into a numerical integration algorithm to produce a piecewise representation of the assessed cumulative distribution, such as those already obtained in the probability and fractile techniques. The dependent measure was taken to be the maximum deviation (vertically) between the piecewise representation of each elicited distribution and the corresponding sample distribution which was shown.

The goodness of fit between the elicited and sample distributions was found to be a nonadditive function of assessment technique and sample distribution shape. Although the fractile procedure performed substantially worse for all three sample distributions, the relative performance of the probability and graph methods varies as a function of sample distribution. The finding that biases in probability assessment result from an interaction between the method of assessment and the shape of the distribution is an important one; the development of an experimental paradigm to adequately evaluate probability assessments is a topic worthy of further attention.

From an applied point of view, this experiment once more suggests that the custom of using fractile techniques for assessing continuous distributions rather than any of the equally simple or simpler alternatives is probably unwise and in need of change. Moreover, it offers evidence for the simplest of all possible alternatives: If you want someone to assess a continuous probability distribution, just ask him to draw it.

II. B. Multiattribute utility analysis: Validation and Application

Four distinct approaches to the validation of multiattribute utilities can be identified.

1. Like preferences, utilities are inherently correct and do not need to be validated.

We, as psychologists knowing that all self-reports can error, consider this idea untenable and not worthy of serious discussion. For that reason, the fact that (as we see it) this view dominates the decision theoretical literature is continually baffling to us.

2. Utilities express strengths of preference; and therefore one validates them by discovering whether or not they correctly predict preference.

If preferences are the ultimate criterion, why bother with utilities? Preferences can be observed directly, and the whole structure of deterministic utility theory is then irrelevant.

3. Utilities are hypothetical constructs, approachable in a number of different ways. Convergent validity is all that can or should be sought. That is, various ways of eliciting values should lead to intelligibly related though not necessarily identical results.

This is an intellectually respectable view, on which we have been doing a lot of work, summarized below.

4. One cannot validate utilities themselves; one can only validate methods of measuring them. This is done by finding or creating stimuli for which values are known, eliciting utilities by various methods, and concluding that the method that most closely approximates the "true" utility is thereby validated. (Note that approaches 3 and 4, though different, do not conflict.)

We remain intellectually much stimulated by this fourth approach, and are beginning to find ways of implementing it. We hope it will be a major theme next year. We think we may see a way of combining it with the third via an application of the Brunswickian lens model approach. We have identified two situations (diamonds and credit risks) in which externally specified and quite explicit multiattribute utility structures already exist.

Convergent validation (to return to approach 3) assumes that a necessary and sufficient condition for a given model or assessment procedure to be valid is that overall utilities which follow from the assessment agree (high Pearson product-moment correlation) with utilities elicited in some other manner. Approaches within this framework, whether "behavioral" or "analytical" in nature, are essentially complex sensitivity analyses in the sense that they study the sensitivity of the output utilities, to such inputs as model structure, elicitation techniques, and respondent identity. Such studies, designed primarily to determine the amount of common variance shared by different modeling or elicitation procedures, are of tremendous practical significance. In particular, the tradeoff between model precision and ease of elicitation must be addressed in almost any application of MAUA.

II. B.1. Model simplifications: A review

Leung (see Summary No. 5) provided a review of theoretical and empirical research findings regarding the sensitivity of MAUA to model specification. The question addressed was whether additional complexities (such as non-additivity, uncertainty, and differential weighting) are useful. Although a few of the studies considered produced analytic solutions to the questions asked, most were either Monte Carlo simulations or behavioral studies. The criterion for intermodel agreement, in almost every study, was the correlation between utilities output by the specified models. Leung came to the following conclusions:

1. Additive models should be used as an approximation to more complicated structures, at least for the two attribute case (unless there are good reasons to believe that a non-additive model is an exact representation of a decision maker's attitudes).

2. Weights do not matter for deterministic additive models when, on the average, the attributes are highly correlated with each other.

3. No conclusions may be drawn regarding how well deterministic models approximated more complicated probabilistic ones.

Of most interest in Leung's analysis was a call for a "measure of robustness other than the coefficient of correlation". Although he was referring to studies involving probabilistic models only, the need for dependent measures of fit is great (see Anderson and Shanteau (1977) for a discussion of this problem).

II. B.2. Monte Carlo simulation: Weighting.

In a study concerning the issue of differential vs. unit weighting for additive deterministic utility functions, Newman (see Summary No. 6) exploited the similarity between the formal mathematical structure of the multiple regression model and the additive utility model (under certainty). Using simulation techniques similar to those described in Newman (1976), Newman (1977) considers two methods of estimating beta weights for regression models and compares them (in terms of variance accounted for on cross validation) to the unit weighting technique. Both procedures for estimating beta weights, ordinary least squares, (OLS) and ridge regression (RIDGE) (Hoerl and Kennard, 1970, a,b) proved superior to unit weighting (UNIT) in all cases save one. In this one case, all the true coefficients were positive, not too far apart, and the sample size was relatively small ($N = 50$). In the overwhelming majority of cases, unit weighting was simply not appropriate.

Newman also found that the ridge estimates outperformed the OLS estimates a great deal of the time, replicating several studies which demonstrate the superiority of the biased RIDGE procedure to the more popular OLS approach (Demster, Schatzoff, and Wermuth, 1975; Hoerl, Kennard, and Baldwin, 1975; Lawless and Wang 1976). Newman asserts that the argument in favor of unit weighting is completely shattered when the differential weigths are estimated via RIDGE. By replacing the independent and criterion

variables in the regression model with the attributes and overall utility construct of a multiattribute utility model, one may ask the following question: What subjective estimation procedure do people use in determining their weights for attributes in a MAUA? The answer to this question is critical. Unit weighting of attributes in a decision analysis is not appropriate if the decision maker can estimate the attribute weights in a RIDGE or even OLS manner. However, if subjective estimates of weights are considerably suboptimal, unit weighting is a boon for the application of MAUA. A study to answer this question, involving a multiple cue probability learning task, is now being planned.

II. B.3. Monte Carlo simulation: Number of attributes.

Leung (see Summary No. 7) described a study to explore the possibility of reducing the number of attributes specified in additive MAU models under certainty. For each example, he randomly generated a utility array (alternative by attribute) and a set of weights for the attributes. Leung systematically varied the number of attributes in the full model, the average intercorrelation among the attributes, the number of attributes in the reduced model, and the method for deciding which attributes to eliminate. Leung investigated the following ad hoc procedures for reducing the number of attributes:

1. Retain the highest weight attribute, drop the attribute that correlates highest with it; repeat until the desired number of attributes are dropped.

2. Ignore intercorrelations, simply drop the desired number of attributes with the lowest weights.

3. Discard the lowest weight attribute: retain the attribute that correlates highest with it; repeat until the desired number of attributes are dropped.

4. Pick the most highly correlated pair of attributes; discard the lower weight attribute of the two; repeat until the desired number of attributes are dropped.

Using as the dependent measure the distribution of correlations (N = 1000) between the full and reduced model, Leung found that methods 2 and 3 (described above) completely dominated methods 1 and 4. He concluded that method 2, considering its ease of application, was the superior procedure. That is, unimportant attributes (attributes which receive small weights) may be eliminated from consideration with little loss. Leung applies

this technique to two real world examples with good results.

The application of the results of this study are subject to the same behavioral questions posed by the Newman study discussed previously. In order to perform Leung's method 2, subjects must be able to accurately rank order attributes in terms of importance. The findings of the proposed multiple cue probability learning study described in II. B.2 would be of obvious importance here.

II. B.4. Behavioral Validation: A construct rather than convergent approach.

In his doctoral dissertation Eils (1977) investigated the use of an external criterion against which to validate additive utility assessments under certainty. Eils elicited utility assessments from twenty-four groups, each of which consisted of four graduate or upper division undergraduate students who knew each other prior to the experimental session. Group utilities were elicited (via consensus) for ten hypothetical applicants for bank credit cards. The research design completely crossed two factors in assessing group utilities: 1) using a decomposition procedure (MAUA) or not and 2) using a formal group communication strategy (GCS) or not. The quality of each group's utility judgments was defined to be the Pearson product-movement correlation between the group's judged utilities and utilities output from a configural (nonlinear) model used by Security Pacific Bank in evaluating applicants for Master Charge. A content analysis of the group's verbal interaction was made to determine the effects of task structure on the characteristics of the group process. Group satisfaction measures were also obtained.

Eils found that the decision technology of MAUA greatly aided groups in reaching decisions that were more consistent (had higher correlations) with decisions based on a systematic collection and interpretation of a large amount of relevant data (i.e. the bank model). When unit weights were used in place of the elicited differential weights, the MAUA groups evidenced even higher correlations with the bank model. The application of a communication strategy did not significantly alter the quality of group evaluations.

Both task interventions (MAUA and GCS) significantly influenced the group communication process. In addition, groups employing the MAUA did

not find the task any more complex or difficult, or any less satisfying than groups not employing the technique. Groups employing GCS did not find their task any less satisfying or complex. Perhaps for the first time, decomposed judgements have been shown to exhibit a greater degree of fit to an external criterion than wholistic judgments. The formalized bank model used to measure judgmental validity reflects the complex nature of the relationship between applicant characteristics and subsequent loan performance. These complex relationships should be similar to the ones inherent in the information that the groups bring to the assessment task in the form of past experience. Thus, the degree to which group decisions correspond to the bank's systematic and complex evaluation provides a measure of how well the elicitation technique taps the information actually contained in group members' past experience. Eils argues that the MAUA procedure he employed proved more valid in that a more complete representation of each individual's past experience was elicited.

II. B.5. Behavioral Validation: Assessment procedures, model structure, separability of attributes, and gains vs. losses.

Eustace and Edwards have designed a study to systematically vary three factors relating to assessment and modeling technique and two factors which describe the nature of the multiattributed entities to be evaluated. In a completely "within" design, Eustace and Edwards elicit two attribute utility functions (either additive or multiplicative) using three popular elicitation techniques (BRTS, Rating Scales, and Certainty Equivalents) and wholistic choices. Another factor, nested withing elicitation technique, is that of risky vs. riskless utility functions. Functions are elicited for gains and losses from starting positions in commodity bundles which are either separable (amounts of tea and ice cream) or inseparable (amount of leanness of ground beef). All assessments are made twice and, with the exception of the wholistic assessments, no real transactions occur between subject and experimenter.

The results from this study will address the following questions:

1. How well do simple utility models (riskless, additive) approximate more complex ones (risky, multiplicative)?

2. How well do simple elicitation procedures (rating scale

holistic choices) approximate more complex ones (BRLTS, Certainty Equivalents)?

3. What is the relationship between gains and losses in starting position when the final utility attained is held in constant?

4. How are the answers to questions 1 through 3 mediated by the separability of attributes in the commodity bundles?

II. B.6. MAUA and Systems Dynamics.

Gardiner and Ford (1977) explain a technique for using additive, riskless, multiattribute utility functions to evaluate the results. Computer simulation models are frequently developed and used as policy analysis tools that show, for the system being modeled, how its behavior over time is influenced by proposed policies. Many simulation efforts stop at this point and leave the synthesis of the derived results to unaided intuitive approaches. The emphasis and focus is on developing models that show consequences of policies, not on the formal evaluation of these consequences. As a result, simulation models and accompanying policy recommendations are frequently criticized for failing to take into account societal interests and values. This paper discussed how MAUA can be applied to the output of computer simulations to remedy the deficiencies inherent in the system dynamics methodology.

The paper discusses an application of the technique in energy boom towns where a system dynamics model of a boom town "feeds" evaluation models developed from nine viewpoints of individuals (including those of the mayor, a conservationist, representatives of the energy industry, etc.) in Framington, New Mexico. The applicability of this technique merger to military boom town phenomena is discussed as well as its application to military "bust towns"; i.e. those instances where U.S. military installations are closed.

III. Applications of Decision Technology

The following section summarizes an ARPA Technical Report by Edwards (1977). Much of that report is concerned with the problems and prospects for institutionalizing decision analytic techniques in Federal bureaucratic contexts, and is in particular responsive to the views on that topic of Mr. Joseph Coates, of the U.S. Office of Technology Assessment. As examples, Edwards summarizes two extensive applications, both using ARPA-developed technology but neither funded by ARPA. Both studies include technical innovations highly relevant to ARPA and DOD needs and problems. The Technical Report makes special effort to be readily understandable by those unacquainted with decision technology, probability, and the like; it does, however, assume experience with Federal bureaucracies.

The first example outlined by Edwards involves the technology of probability assessment and use of Bayes' theorem. Probability of various diagnoses were assessed by clinicians in emergency room settings all over the U.S. to determine the diagnostic efficacy of the radiographic procedures employed. Specifically, clinicians provided probability diagnoses before and after interpretation of about 8,000 x-rays. The log likelihood ratio, computed from the prior and posterior probabilities assessed, served as a measure of the influence of x-rays on clinical diagnosis. The assessments were accomplished "in the field" by clinicians with a minimum of technical training.

The example suggests extensions of the described methodology to a variety of real world settings. Any situation in which a costly, perhaps dangerous, procedure to gather information is employed is amenable to this investigatory approach. As technology in general advances and methods to reduce uncertainty become increasingly more available, the decision of whether the amount of additional information obtained is worth the energy expended in gathering it will become both more important and more complex. As technological sophistication increases, the stakes increase and the intuitive ability of man to choose beneficially between seeking or not seeking more information decreases. Thus, techniques of studying the efficacy (in some sense) of information collecting procedures (such as radiography) will become increasingly important. The most obvious military example has to do with collection of intelligence information.

In Edwards' second example, the technique of multiattribute utility analysis (MAUA) is applied to a highly complex social decision making problem; siting a nuclear waste disposal facility. In contrast to the first example, the primary focus is on determining measures of value, not uncertainty. The most important feature of this application is the use of the MAUA procedure (developed for use by individuals) by a face-to-face group of decision makers.

Group interactions were structured around the MAUA tasks of determining dimensions of importance, and weighting those dimensions. Experts in nuclear engineering from several countries comprised the groups. Hypothetical alternative waste disposal sites were generated by one of the experts who had extensive experience with the siting problem. A numerical demonstration of MAUA evaluation of sites was performed, using the weights assessed from the experts and linear transformations of values (or log values) as location measures on utility curves.

Two additions to the usual MAUA technique were employed. Rather than obtaining ratio scaled weights in which only ratios involving the least important attribute are checked, the respondents were required to judge ratios of all possible pairs of weights. This change in elicitation procedure probably enhanced the reliability, and hence validity, of the utlity model parameter estimates determined via the weighting procedure.

Another unique aspect of this endeavor is the rescaling of weights to reflect the range of values on each dimension for which alternatives are actually available. Since weights must often be obtained using a best guess of the ranges on each dimension which the alternatives will span, this rescaling procedure is a potentially valuable one. Further research exploring the accuracy of the assumptions implicit in this technique is still necessary, however.

IV. REFERENCES

Alpert, M. and Raiffa, H. A progress report on the training of probability assessors. Unpublished manuscript, Harvard Un.1969.

Anderson, N.H. and Shanteau, J. Weak inference with linear models. Psychol. Bull. 84, 1155-1170, 1977.

Arrow, K. Social Choice and Individual Values. New York: Wiley, 1951.

Dalkey, N. An impossibility theorem for group probability functions, Rand Paper P-4862, The Rand Corporation, Santa Monica, California, 1972.

Dalkey, N. and Helmer, O. An experimental application of the Delphi Method to the use of experts. Management Science , 1963, 9, 458-467.

DeGroot, M. Reaching a consensus, Journal of American Statistical Association. 1974, 69 , 118-121.

Delbecq, A. and Van de Ven, A. A group process model for problem identification and program planning, Journal of Applied Behavioral Science, 1971, 7, 466-492.

Dempster, A.P., Schatzoff, M., Wermuth, N.A. Simulation study of alternatives to ordinary least squares. Research Report S-35, Department of Statistics, Harvard University, 1976.

Edwards, W. Technology for director dubious: Evaluation and decision in public contexts, Technical Report 77-2, Social Science Research Institute, August, 1977.

Edwards, W. Research on the technology of inference and decision. Final Technical Report, 75-10, Social Science Research Institute, August, 1975.

Edwards, W. Research on the technology of inference and decision, Final Technical Report, 011313-F, Social Science Research Institute, October, 1973.

Edwards, W., Phillips, L.D., Hays, W.L., and Goodman, B.C. Probabilistic information processing systems: Design and evaluation. IEEE Transactions on Systems Science and Cybernetics, 1968. SSC-4, 248-1120.

Edwards, W. and Seaver,D.A. Research on the technology of inference and decision. Final Technical Report. 76-7, Social Science Research Institute, October, 1976.

Eils, L.C. III, Seaver, D.A., and Edwards, W. Developing the technology of probabilistic inference: Aggregating by averaging reduces conservatism. Technical Report 77-3. Social Science Research Institute, August, 1977.

Eils, L.C. III. The effects of communication strategy and decision technology on the process and product of decision making groups. Unpublished doctoral dissertation, University of Southern California, June, 1977.

Gardiner, P., and Ford, A. In Snapper (ed) Models in Metrics for Decision Makers. Washington, D.C. : Information Resources Press, (in press).

Goodman, B. Direct estimation procedures for eliciting judgments about uncertain events. Technical Report No. 011313-5-T, Engineering Psychology Laboratory, University of Michigan, 1973.

Hoerl, A.E. and Kennard, R.W. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 1970, 12, 55-67.

Hoerl, A. E. and Kennard, R.W. Ridge regression: Applications to nonorthogonal problems. Technometrics, 1970, 12, 69-82.

Hoerl, A. E. Kennard, R.W., and Baldwin, D.F. Ridge regression: Some simulations. Communications in Statistics, 1975, 4, 105-123.

John, R. and Edwards, W. Estimating subjective probability distributions, Working Paper, Social Science Research Institute, 1977.

Lawless, J. F. and Wang, P. A simulation study of ridge and other regression estimators. Communications in Statistics, 1976, 5, 307-323.

Phillips, L., and Edwards, W. Conservatism in a simple probability inference task. Journal of Experimental Psychology. 1966, 72, 346-352.

Rabin, J. B. Ready reserve account/master charge credit scoring study. Technical Report, Marketing Department, Management Services Division, Security Pacific Bank, 1974.

Seaver, D., von Winterfeldt, D., and Edwards W. Eliciting subjective probabilities on continuous variables. Research Report No. 75-8, Social Science Research Institute, University of Southern California, August, 1975.

Shanteau, J. Correlation as a deceiving measure of fit. Bulletin of the Psychonomic Society, 1977, 10, 134-136.

Slovic, P., Fischoff, B., Lichtenstein, S., Combs, B., and Layman, M. Perceived frequency of low probability, lethal events. Unpublished manuscript, 1975.

Stillwell, W.G., Seaver, D.A., and Edwards, W. The effects of response scales on likelihood ratio judgments. Technical Report 77-5, Social Science Research Institute, August, 1977.

Winkler, R.L. The quantification of judgment: Some methodological suggestions. Journal of the American Statistical Association, 1967. 62, 1105-1120.

V. Summaries of Technical Reports

Summary No. 1

How Groups can Assess Uncertainty: Human Interaction Versus Mathematical Models

David A. Seaver

Recently developed decision aiding technologies rely upon quantification of uncertainty as subjective probability. Since groups are often responsible for making decisions, procedures for assessing the subjective probabilities of groups are necessary if decision analytic techniques are to be generally applicable. Two general approaches to this problem exist: mathematical aggregation, in which individual probabilities are combined via some mathematical rule to from a single probability assessment, and behavioral interaction, in which the group members communicate verbally or otherwise to reduce or eliminate disagreement. Several methods in each of these categories are reviewed. Since previous results comparing various procedures for determining group probabilities are equivocal, a study was undertaken to compare several mathematical aggregation and behavioral interaction approaches. The results of this study suggested that some interaction tends to increase the certainty of the group, decrease the calibration, and decrease the disagreement among group members, although the type of interaction makes little difference. The mathematical aggregation rule used affects both the calibration and the certainty of the group. Choice of just which procedure to use depends on a tradeoff between the desirability of increased certainty and calibration. In many instances, simple averaging of individual assessments without any group interaction may be the most desirable procedure simply because it is the easiest to use.

Summary No. 2

The Effects of Response Scales on Likelihood Ratio Judgments

William G. Stillwell, David A. Seaver, and Ward Edwards

Different methods of eliciting responses to the same question often produce different responses. In order to systematically study how response scales affect likelihood ratio judgments, two experiments were conducted. Experiment I manipulated two independent variables: the endpoints of the response scales (100:1, 1000:1, 10,000:1) and the spacing of the scales (logarithmic versus linear). Results compared the veridicality of responses on the six scales produced by crossing these factors plus another response mode in which subjects simply wrote their judgment in a blank (no scale).

Logarithmic scales produced responses that were both more veridical and more consistent than responses on linear scales which were, in turn, better than simple written responses. Measures of the effects of the endpoints were somewhat inconsistent and probably interacted with the range of veridical likelihood ratios. Judgments of relatively small likelihood ratios were affected by the spacing: linear spacing caused overestimation. Judgments of relatively large likelihood ratios were controlled more by the endpoints: higher endpoints produced larger judgments. Apparently, subjects use the range of the scale as information about the range of true likelihood ratios.

Experiment II manipulated two additional variables, data diagnosticity and the values of the true likelihood ratios. The results of Experiment I were confirmed while neither of the additional variables radically changed the effects of endpoints or spacing.

Summary No.3

Developing the Technology of Probablistic Inference
Aggregating by Averaging Reduces Conservatism

Lee C. Eils, III, David A. Seaver, and Ward Edwards

A relatively large body of research indicates that people are conservative processors of probabilistic information. Recent attention has focused on two possible explanations of this phenomenon. The misaggregation hypothesis depicts convservatism as an inability to properly combine the information in a data sequence. The other explanation suggests conservatism is the result of a response bias: the avoidance of extreme odds or probability judgments.

Two experiments explored the use of a specific response, average certainty, that was devised to thwart conservatism caused by either response bias or misaggregation. Use of appropriate instructions and response scales made the average certainty judgments good subjective assessments of the arithmetic mean likelihood ratio which could then be used in the appropriate form of Bayes' Theorem to calculate posterior odds. These judgments seemed unlikely to be affected by a response bias since extreme responses were not needed. In addition, research has suggested that people are more likely to aggregate information by averaging than by adding or multiplying, so misaggregation may be exhibited only in specific forms of aggregation and may not be present in averaging.

The results of Experiment I indicated that average certainty judgments were both more orderly and more veridical than cumulative certainty judgments of the type usually obtained in probabilistic inference tasks. The cumulative judgments were very conservative while the average certainty judgments were only slightly radical. Experiment II indicated that average certainty judgments and individual likelihood ratio judgments were both more orderly and veridical than cumulative certainty judgments but that they did not differ significantly from each other in either orderliness or veridicality. A second factor, the diagnosticity level of the data was also found to influence the veridicality of obtained judgments. Regardless of the method

of aggregation employed, estimates became more veridical as the data became more diagnostic. Since thse studies were undertaken only to see if average certainty judgments are an effective way to reduce conservatism, they do not directly test what causes conservatism. However, some implications concerning the nature of conservatism are discussed, as are the implications for the technology of probabilistic inference.

Summary No. 4

Subjective Probability Assessment and the Shape of the Distribution

Richard S. John and Ward Edwards

Seventy-two subjects were presented with three samples of pickup sticks, each painted yellow and blue. After viewing each sample distribution, subjects assessed subjective probability distributions over the "length of yellow" painted on sticks in each population sampled. Each subject utilized one of three popular probability assessment techniques in making the uncertainty judgments. The shape of the distribution of lengths of yellow was found to interact with assessment technique, suggesting that biases introduced in subjective probability distributions vary as a function of the uncertain quantity being assessed. The customary "fractile" procedure for assessing continuous probability distributions consistenty yielded the worst fitting subjective assessments.

Summary No. 5

Sensitivity Analysis of the Effect of Variation in the Form and Parameters of a Multiattribute Utility Function: A Survey

Patrick Leung

There is a trend towards the development of complicated versions of multiattribute utility models. These models, although theoretically more accurate in the representation of decision makers' attitudes, require assessment procedures which are more difficult and time consuming to implement than simpler models. The paper reviews theoretical and empirical research on the sensitivity of multiattribute utility models with emphasis on simplification. Both deterministic and probabilistic models are considered and the studies are divided into four areas: 1) those involving sensitivity to the form of the multiattribute utility function; 2) those involving sensitivity to the parameters of the function; 3) those involving sensitivity to the form of individual single attribute utility functions; and 4) those involving the relationship between deterministic and probabilistic models.

Summary No. 6

Differential Weighting for Prediction and Decision Making Studies

A Study of Ridge Regression

J. Robert Newman

This paper is another in a series exploring the conditions under which either differential or simple unit weighting of predictor variables in prediction and/or decision studies will be appropriate. Some of the difficulties of applying the ordinary least squares (OLS) analysis to practical problems are described and an alternative regression model called ridge analysis (RIDGE) is offered as a substitute to OLS. The trouble with OLS is that when the predictor variables are intercorrelated, then the regression coefficients estimated by OLS are often quite deviant from the "true" coefficients. They are often too large in absolute value and the sign of the coefficient can be wrong. The RIDGE solution to this is very simple: just add small positive values to the main diagonal of the correlation matrix depicting the intercorrelations between the predictor variables, and re-estimate the coefficients in the usual manner. The resulting estimates are called RIDGE estimates and in theory they will be superior to OLS estimates in the sense of producing smaller error in cross validation samples. That is, when OLS and RIDGE estimates are estimated in one sample of data, and then tested on a new sample of data, the RIDGE estimates will result in fewer errors of prediction than the OLS estimates.

Several empirical studies were conducted using computer simulated data for various prediction situations. The OLS and RIDGE models were compared as to their efficacy in prediction and both models were compared against the simplest model possible, that of unit weighting (UNIT), in which no weighting is performed; the variables are simply added up and the sum used for prediction. The results of these studies indicate that OLS and RIDGE, with one exception, always outperformed UNIT with respect to producing smaller errors of prediction and, what is more important, RIDGE always did better than OLS. The one exception

in which UNIT did better than OLS and RIDGE is for the case in which all the "true" coefficients are positive, not too far apart, and the sample size is relatively small (≤ 50). This is a very restricted class of conditions. The general conclusion is that UNIT weighting will be preferred as a way of generating differential weights. Also the RIDGE method of estimation (RIDGE) always should be the preferred model over OLS. One practical implication of this is that if an investigator does not have the luxury to do cross validation then RIDGE estimation can be used as a substitute for cross validation.

Summary No. 7

The Effects of Reducing the Number of Attributes in Additive Multiattribute Utility Modeling Under Certainty

Patrick Leung

This paper explores the effects of reducing the number of attributes in multiattribute utility modeling under certainty. Four different schemes for reducing the number of attributes are tested, using Monte Carlo Simulation. A simple method which ignores intercorrelations among attributes and takes only the weights into account is found to yield a reduced model whose correlation with the original full model is highest. This method is applied to two real world examples--an automobile evaluation problem and a coastal development site selection problem- and yields good results in both cases. The amount of time and effort saved through the use of the reduced model instead of the full model is found to be considerable.

CONTRACT DISTRIBUTION LIST
(Unclassified Technical Reports)

Director Advanced Research Projects Agency Attention: Program Management Office 1400 Wilson Boulevard Arlington, Virginia 22209	2 copies
Office of Naval Research Attention: Code 455 800 North Quincy Street Arlington, Virginia 22217	3 copies
Defense Documentation Center Attention: DDC-TC Cameron Station Alexandria, Virginia 22314	12 copies
DCASMA Baltimore Office Attention: Mr. K. Gerasim 300 East Joppa Road Towson, Maryland 21204	1 copy
Director Naval Research Laboratory Attention: Code 2627 Washington, D.C. 20375	6 copies
Office of Naval Research Attention: Code 102IP 800 North Quincy Street Arlington, Virginia 22217	6 copies
Decisions and Designs, Incorporated 8400 Westpark Drive, Suite 600 McLean, Virginia 22101	20 copies

Revised December 1977

SUPPLEMENTAL DISTRIBUTION LIST
(Unclassified Technical Reports)

Department of Defense

Director of Net Assessment
Office of the Secretary of Defense
Attention: MAJ Robert G. Gough, USAF
The Pentagon, Room 3A930
Washington, DC 20301

Assistant Director (Net Technical Assessment)
Office of the Deputy Director of Defense
Research and Engineering (Test and Evaluation)
The Pentagon, Room 3C125
Washington, DC 20301

Assistant Director (Environmental and Life Sciences)
Office of the Deputy Director of Defense
Research and Engineering (Research and Advanced Technology)
Attention: COL Henry L. Taylor
The Pentagon, Room 3D129
Washington, DC 20301

Director, Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

Director, Cybernetics Technology Office
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

Director, ARPA Regional Office (Europe)
Headquarters, U.S. European Command
APO New York 09128

Director, ARPA Regional Office (Pacific)
Staff CINCPAC, Box 13
APO San Francisco 96610

Dr. Don Hirta
Defense Systems Management School
Building 202
Ft. Belvoir, VA 22060

Chairman, Department of Curriculum Development
National War College
Ft. McNair, 4th and P Streets, SW
Washington, DC 20319

Defense Intelligence School
Attention: Professor Douglas E. Hunter
Washington, DC 20374

Vice Director for Production
Management Office (Special Actions)
Defense Intelligence Agency
Room 1E863, The Pentagon
Washington, DC 20301

Command and Control Technical Center
Defense Communications Agency
Attention: Mr. John D. Hwang
Washington, DC 20301

Department of the Navy

Office of the Chief of Naval Operations (OP-951)
Washington, DC 20450

Office of Naval Research
Assistant Chief for Technology (Code 200)
800 N. Quincy Street
Arlington, VA 22217

Office of Naval Research (Code 230)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Naval Analysis Programs (Code 431)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Operations Research Programs (Code 434)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Information Systems Program (Code 437)
800 North Quincy Street
Arlington, VA 22217

Director, ONR Branch Office
Attention: Dr. Charles Davis
536 South Clark Street
Chicago, IL 60605

Director, ONR Branch Office
Attention: Dr. J. Lester
495 Summer Street
Boston, MA 02210

Director, ONR Branch Office
Attention: Dr. E. Gloye
1030 East Green Street
Pasadena, CA 91106

Director, ONR Branch Office
Attention: Mr. R. Lawson
1030 East Green Street
Pasadena, CA 91106

Office of Naval Research
Scientific Liaison Group
Attention: Dr. M. Bertin
American Embassy - Room A-407
APO San Francisco 96503

Dr. A. L. Slafkosky
Scientific Advisor
Commandant of the Marine Corps (Code RD-1)
Washington, DC 20380

Headquarters, Naval Material Command
(Code 0331)
Attention: Dr. Heber G. Moore
Washington, DC 20360

Head, Human Factors Division
Naval Electronics Laboratory Center
Attention: Mr. Richard Coburn
San Diego, CA 92152

Dean of Research Administration
Naval Postgraduate School
Attention: Patrick C. Parker
Monterey, CA 93940

Naval Personnel Research and Development
Center (Code 305)
Attention: LCDR O'Bar
San Diego, CA 92152

Navy Personnel Research and Development
Center
Manned Systems Design (Code 311)
Attention: Dr. Fred Muckler
San Diego, CA 92152

Naval Training Equipment Center
Human Factors Department (Code N215)
Orlando, FL 32813

Naval Training Equipment Center
Training Analysis and Evaluation Group
(Code N-00T)
Attention: Dr. Alfred F. Smode
Orlando, FL 32813

Director, Center for Advanced Research
Naval War College
Attention: Professor C. Lewis
Newport, RI 02840

Naval Research Laboratory
Communications Sciences Division (Code 540
Attention: Dr. John Shore
Washington, DC 20375

Dean of the Academic Departments
U.S. Naval Academy
Annapolis, MD 21402

Chief, Intelligence Division
Marine Corps Development Center
Quantico, VA 22134

Department of the Army

Alan H. Curry
Operations and Management Science Division
U.S. Army Institute for Research in Manage
ment Information and Computer Science
730 Peachtree St., N.E. (Suite 900)
Atlanta, Georgia 30308

Other Government Agencies

Chief, Strategic Evaluation Center
Central Intelligence Agency
Headquarters, Room 2G24
Washington, DC 20505

Director, Center for the Study of Intelligence
Central Intelligence Agency
Attention: Mr. Dean Moor
Washington, DC 20505

Mr. Richard Heuer
Methods & Forecasting Division
Office of Regional and Political Analysis
Central Intelligence Agency
Washington, DC 20505

Office of Life Sciences
Headquarters, National Aeronautics and Space Administration
Attention: Dr. Stanley Deutsch
600 Independence Avenue
Washington, DC 20546

Other Institutions

Department of Psychology
The Johns Hopkins University
Attention: Dr. Alphonse Chapanis
Charles and 34th Streets
Baltimore, MD 21218

Institute for Defense Analyses
Attention: Dr. Jesse Orlansky
400 Army Navy Drive
Arlington, VA 22202

Director, Social Science Research Institute
University of Southern California
Attention: Dr. Ward Edwards
Los Angeles, CA 90007

Perceptronics, Incorporated
Attention: Dr. Amos Freedy
6271 Variel Avenue
Woodland Hills, CA 91364

Stanford University
Attention: Dr. R. A. Howard
Stanford, CA 94305

Director, Applied Psychology Unit
Medical Research Council
Attention: Dr. A. D. Baddeley
15 Chaucer Road
Cambridge, CB 2EF
England

Department of Psychology
Brunel University
Attention: Dr. Lawrence D. Phillips
Uxbridge, Middlesex UB8 3PH
England

Decision Analysis Group
Stanford Research Institute
Attention: Dr. Miley W. Merkhofer
Menlo Park, CA 94025

Decision Research
1201 Oak Street
Eugene, OR 97401

Department of Psychology
University of Washington
Attention: Dr. Lee Roy Beach
Seattle, WA 98195

Department of Electrical and Computer Engineering
University of Michigan
Attention: Professor Kan Chen
Ann Arbor, MI 94135

Department of Government and Politics
University of Maryland
Attention: Dr. Davis B. Bobrow
College Park, MD 20747

Department of Psychology
Hebrew University
Attention: Dr. Amos Tversky
Jerusalem, Israel

Dr. Andrew P. Sage
School of Engineering and Applied Science
University of Virginia
Charlottesville, VA 22901

Professor Raymond Tanter
Political Science Department
The University of Michigan
Ann Arbor, MI 48109

Deputy Under Secretary of the Army
(Operations Research)
The Pentagon, Room 2E621
Washington, DC 20310

Director, Army Library
Army Studies (ASDIRS)
The Pentagon, Room 1A534
Washington, DC 20310

U.S. Army Research Institute
Organizations and Systems Research Laboratory
Attention: Dr. Edgar M. Johnson
5001 Eisenhower Avenue
Alexandria, VA 22333

Director, Organizations and Systems
Research Laboratory
U.S. Army Institute for the Behavioral
and Social Sciences
1300 Wilson Boulevard
Arlington, VA 22209

Technical Director, U.S. Army Concepts
Analysis Agency
8120 Woodmont Avenue
Bethesda, MD 20014

Director, Strategic Studies Institute
U.S. Army Combat Developments Command
Carlisle Barracks, PA 17013

Commandant, Army Logistics Management Center
Attention: DRXMC-LS-SCAD (ORSA)
Ft. Lee, VA 23801

Department of Engineering
United States Military Academy
Attention: COL A. F. Grum
West Point, NY 10996

Commanding General
Headquarters, DARCOM
Attention: DRCED - Richard Murray
5001 Eisenhower Avenue
Alexandria, VA 22333

Marine Corps Representative
U.S. Army War College
Carlisle Barracks, PA 17013

Chief, Studies and Analysis Office
Headquarters, Army Training and Doctrine
Command
Ft. Monroe, VA 23351

Commander, U.S. Army Research Office
(Durham)
Box CM, Duke Station
Durham, NC 27706

Department of the Air Force

Assistant for Requirements Development
and Acquisition Programs
Office of the Deputy Chief of Staff for
Research and Development
The Pentagon, Room 4C331
Washington, DC 20330

Air Force Office of Scientific Research
Life Sciences Directorate
Building 410, Bolling AFB
Washington, DC 20332

Commandant, Air University
Maxwell AFB, AL 36112

Chief, Systems Effectiveness Branch
Human Engineering Division
Attention: Dr. Donald A. Topmiller
Wright-Patterson AFB, OH 45433

Director, Advanced Systems Division
(AFHRL/AS)
Attention: Dr. Gordon Eckstrand
Wright-Patterson AFB, OH 45433

Commander, Rome Air Development Center
Attention: Mr. John Atkinson
Griffis AFB
Rome, NY 13440

IRD, Rome Air Development Center
Attention: Mr. Frederic A. Dion
Griffis AFB
Rome, NY 13440

HQS Tactical Air Command
Attention: LTCOL David Dianich
Langley AFB, VA 23665

Professor Howard Raiffa
Morgan 302
Harvard Business School
Harvard University
Cambridge, MA 02163

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 01855-6-T
2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle): Research on the Technology of Inference and Decision

5. TYPE OF REPORT & PERIOD COVERED: Final Technical Report 10/1/76-9/30/77

6. PERFORMING ORG. REPORT NUMBER: None

7. AUTHOR(s): Ward Edwards, Richard John, and William Stillwell

8. CONTRACT OR GRANT NUMBER(s): Prime Contract N00014-76-C-0074; Subcontract 76-030-0715

9. PERFORMING ORGANIZATION NAME AND ADDRESS: Social Science Research Institute, University of Southern California, Los Angeles, California 90007

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS: Advanced Research Projects Agency, 1400 Wilson Blvd., Arlington, Virginia 22209

12. REPORT DATE: November 1977

13. NUMBER OF PAGES: 37

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Decisions and Designs, Inc., Suite 100, 7900 Westpark Drive, McLean, Virginia 22101 (under contract from Office of Naval Research)

15. SECURITY CLASS. (of this report): Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

multiattribute utility; log likelihood ratio; simulation; validation; elicitation procedures; response modes; small probabilities; biases; Ridge regression; least squares regression; weighting; attributes

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report summarizes twelve months of research on the technology of inference and decision. Theoretical research and experimental work on three major topics; elicitation of subjective probabilities, multi-attribute utility theory, and the application of decision technology, is discussed. Experimental work showed that simple averaging of individual's probability judgments to form a group judgment did not differ significantly from behavioral interaction in final quality of the judgments as evaluated

by a quadratic scoring rule. Other experimental work indicated that elicitation techniques were of significant importance to the quality of judgments. response scales were found to affect both the magnitude and veridicality of probabilistic judgment. In the assessment of subjective probability distributions elicitation technique was found to interact with the type of distribution used to generate the data in that biases introduced in subjective probability distributions varied as a function of the uncertain quantity being assessed.

Simplification techniques for the assessment of multiattributed utilities were investigated and it was found that several methods for the selection of subsets of the total number of attributes lead to remarkably robust results. Ridge regression was tested as an alternative to the standard least squares method of estimating weights in the assessment of utility under certainty and found to outperform the least squares procedure a great deal of the time. The marriage of multiattribute utility assessment (MAUA) and systems dynamics was undertaken with the MAUA used as evaluation for the outputs of the dynamic model.

A practical discussion of the application of decision technology examined its usefullness in Department of Defense contexts. Problems with the institutionalizing of decision analysis techniques in Federal bureaucratic contexts were discussed as well as two applications illustrating procedures.